

STRUCTURE OF THE HUMAN CR1 GENE

Molecular Basis of the Structural and Quantitative Polymorphisms and Identification of a New CR1-like Allele

WINNIE W. WONG,*† JEANNE M. CAHILL,* MATTHEW D. ROSEN,*
CHRISTINE A. KENNEDY,* ERMELINDA T. BONACCIO,*
MICHAEL J. MORRIS,* JAMES G. WILSON,*† LLOYD B. KLINKSTEIN,§
AND DOUGLAS T. FEARON||

*From the *Department of Rheumatology and Immunology, Brigham and Women's Hospital; the †Department of Medicine and the §Program in Cell and Developmental Biology, Harvard Medical School, Boston, Massachusetts 02115; and the ||Department of Medicine, The Johns Hopkins University School of Medicine and The Johns Hopkins Hospital, Baltimore, Maryland 21205*

The human receptor for the complement cleavage fragments C3b and C4b (complement receptor type 1, CR1) is found on the surfaces of erythrocytes, most peripheral blood leukocytes, glomerular podocytes, and follicular dendritic cells, and in plasma (1-3). Besides the removal of C3b- or C4b-coated microorganisms or immune complexes (4), CR1 also serves as an inhibitor of the C3 and the C5 convertases by dissociating C2b and Bb fragments and by acting as a cofactor for the factor I-mediated cleavage of C3b and C4b (5, 6). This regulatory capacity is shared by the structurally related members of a supergene family, termed the regulator of complement activation (RCA)¹ region, located on chromosome 1 (7-13).

Human CR1 is a single chain glycoprotein with four allotypic variants that differ in M_r on SDS-PAGE by increments of 40-50 kD (14-17). The two most common variants are termed F and S (or A and B allotypes) and exhibit M_r of 250 and 290 kD, respectively. These variations reflect differing lengths of the polypeptides and not posttranslational modifications, since distinct unglycosylated precursors have been described (18) and incremental differences of 1.3 to 1.5 kb were also observed in the CR1 transcripts from various allotypes (19, 20). The cDNA that encodes the entire F allotype of 2,039 amino acids (aa) has been sequenced and the extracellular portion of the molecule is found to comprise 30 short consensus repeats (SCR). A feature that distinguishes CR1 from the other proteins of this gene family is the organization of the NH₂-terminal 28 SCRs into four tandem long homologous repeats (LHR) of about 450 aa, each containing seven SCRs. Extensive sequence homologies of 60 to 99% have been observed among the LHRs, suggesting that they have arisen by gene duplication (21, 22).

This work was supported by National Institutes of Health grants AR-37408, AI-22833, and AI-23401 and by the Arthritis Foundation Biomedical Research Center Grant. J. G. Wilson's current address is Veteran's Administration Medical Center, 1500 Woodrow Wilson Dr., Jackson, MS 39216. Address correspondence to W. Wong, Room 508, Seeley G. Mudd Building, 250 Longwood Ave., Boston, MA 02115.

¹ *Abbreviations used in this paper:* aa, amino acid; LHR, long homologous repeat; RCA, regulator of complement activation; SCR, short consensus repeat; RFLP, restriction fragment length polymorphism.

The genetic mechanisms that led to the generation of the allotypic forms of CR1 were initially explored by the analyses of the RFLP of the DNA from donors expressing the F or the S allotypes (19). The evidence against the generation of the differently sized allotypes by alternative splicing of the transcripts are threefold. First, the appearance of additional Bam HI and Sac I fragments were associated with the S allotype without the corresponding loss of any fragments in the DNA of homozygous FF donors, indicating that insertion or deletion events, rather than mechanisms involving single base changes, were responsible for these RFLPs. Second, introns were duplicated in the F allele and an additional duplication was present in the S allele. Third, the allotypic differences of 40–50 kD at the protein level and 1.4 kb at the transcriptional level corresponded to the length of one LHR, suggesting that addition or deletion of this unit formed the basis of the structural polymorphism in human CR1 (19).

A quantitative polymorphism has also been observed in CR1 in which the number of receptors on erythrocytes was determined by an autosomal codominant mode of inheritance (23–26). High or low CR1 expression on erythrocytes was associated with allelic 7.4- and 6.9-kb genomic Hind III fragments, respectively (25, 26). Linkage disequilibrium was observed between this RFLP and the CR1 structural allotypes, suggesting that this polymorphism was located within or near the CR1 gene (19, 24, 25). The present study examines the molecular bases of these structural and genomic polymorphisms by mapping the CR1 gene and localizing the restriction sites that determine these RFLPs.

Materials and Methods

Genomic Libraries. Large fragments of the DNA from peripheral blood leukocytes or EBV-transformed cell lines of selected donors were obtained by partial digestion with Sau 3A I and selection on a 10–40% sucrose gradient. The fractions containing DNA of 35–45 kb were pooled, precipitated in ethanol, and ligated to the cosmid vectors pJB8 or pNNRI (27). The plasmid pNNRI was derived from pTCF (28) by substitution of the Hpa I site by a Not I site. For the construction of genomic libraries in phage, the DNA fragments of 18–25 kb from the sucrose gradient were ligated to EMBL-3 arms (29) (Stratagene, San Diego, CA) and the recombinant clones were amplified on *Escherichia coli* strain P2392. The genomic library in EMBL-3 of a donor who expressed the S allotype of CR1 was the gift of Dr. John H. Weis (Harvard Medical School, Boston, MA) (21). Recombinant CR1 clones were identified by hybridization to ³²P-labeled cDNA probes.

Southern Blot Analysis and Mapping of Recombinant Clones. DNA was isolated from peripheral blood leukocytes or EBV-transformed cell lines and analyzed after digestion with restriction enzymes and agarose gel electrophoresis as previously described (19). The DNA from the transformed cell lines was compared on Southern blots to DNA isolated from the peripheral blood leukocytes of the same donor to ascertain that rearrangement had not occurred. Recombinant phage or cosmid DNA was isolated by centrifugation on CsCl gradients (27) and analyzed by probing the blots of the restriction digests with cDNA or genomic fragments. The order of the restriction sites was determined by probing the blots of the partial digests with DNA fragments derived from the left or right arm of phage λ , termed MAP-L and MAP-R, respectively (30). The restriction maps of the cosmid clones were similarly determined by analysis of Not I-linearized DNA using probes specific for the fragments to the right or to the left of the Hpa I/Not I site.

Probes. Restriction fragments of CR1 cDNA (Fig. 1 A) were subcloned into PUC18. Owing to the high sequence homology among the LHRs, crosshybridization is observed (Table I). Thus, although CR1-1 is derived from LHR-B, it hybridizes to corresponding regions in LHR-A, -C, and -D. CR1-4 hybridizes to the first two SCRs in LHR-B and -C, while CR1-18

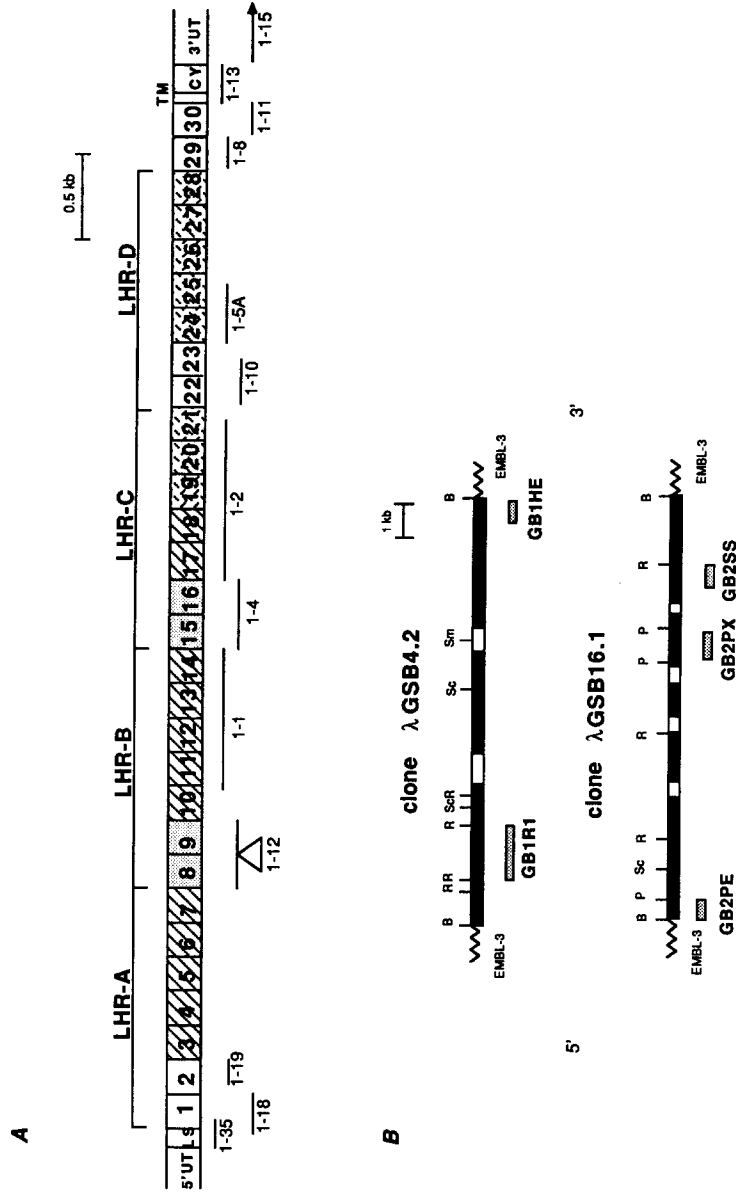


FIGURE 1. Complementary DNA and genomic probes used in the mapping of the CR1 gene. (A) The functional domains encoded by the cDNA of the F allotype of CR1 are represented by the boxes marked as follows: UT, untranslated; LS, leader sequence; 1-30, the SCRs of ~60 amino acids each; TM, transmembrane region; and CY, cytoplasmic domain. The first 28 SCRs are organized into four LHRs designated by the brackets marked LHR-A, -B, -C, and -D. The regions of >90% sequence homology are shaded by the same patterns, while unique sequences are left blank. The lines represent the cDNA probes derived from the corresponding regions and these are designated by numbers only in subsequent figures. CR1-12 is a part of an intron between SCR-8 and -9 in LHR-B (21). (B) The black bars represent the two polymorphic Bam HI fragments that comprise the inserts of clones λ GSB4.2 and λ GSB16.1. The white boxes represent the regions that hybridize to the cDNA probes CR1-1 or CR1-4 and the stippled boxes represent the genomic probes derived from the corresponding segments of DNA. The restriction sites are designated as follows: B, Bam HI; R, Eco RI; P, Pvu II; Sc, Sac I and Sm, Sma I.

TABLE I.
Patterns of Hybridization of CR1 cDNA Probes

LHR-A		LHR-B		LHR-C		LHR-D	
SCR	Probe	SCR	Probe	SCR	Probe	SCR	Probe
1	CR1-18	8	CR1-4	15	CR1-4	22	CR1-10
2	CR1-19	9	CR1-4	16	CR1-4	23	CR1-10
3	CR1-2	10	CR1-2	17	CR1-2	24	CR1-5A
4	CR1-1,1-2	11	CR1-1,1-2	18	CR1-1,1-2	25	CR1-5A
5	CR1-1,1-2	12	CR1-1,1-2	19	CR1-1,1-2	26	CR1-1,1-2
6	CR1-1,1-2	13	CR1-1,1-2	20	CR1-1,1-2	27	CR1-1,1-2
7	CR1-1	14	CR1-1	21	CR1-1,1-2	28	CR1-1,1-2

and CR1-19 are specific for the 5' region of LHR-A and CR1-10 for the 5' region of LHR-D (21, 22). CR1-12 represents a segment of an intron between SCR-8 and -9 (21).

The intervening sequence probes, GB1R1, GBIHE (HE), GB2PE (PE), GB2PX (PX), and GB2SS (SS) were derived from the two distinct polymorphic Bam HI fragments that had been cloned from selected donors (19) (Fig. 1 B). The insert from clone 16.1 is associated exclusively with the expression of the S allotype and is expected to lie within the structural gene of CR1 (19). GB2PE, -PX, and -SS are Pvu II-Bam HI, Pvu II-Xba I and Ssp I fragments of ~400, ~800, and ~300 bp, respectively. GB1R1 and -HE are Eco RI and Hind III-Bam HI fragments of 1.85 kb and ~400 bp, respectively. Probe 2.22.2, derived from the 3' end of phage clone 2.22, is located in a 15-kb intron in LHR-D. None of the genomic probes contained CR1-specific coding sequences of human repeat sequences.

DNA Sequence Analysis. Restriction fragments of genomic clones were subcloned into M13mp18 or M13mp19 and sequenced by the dideoxynucleotide chain termination method (31) using a Sequenase Kit (U. S. Biochemical, Cleveland, OH) (32). DNA sequence analyses were performed at the computer facility of the Howard Hughes Medical Institute, Harvard Medical School Department of Genetics, using the Genetics Computer Group software of the University of Wisconsin Biotechnology Center (33).

Results

The Organization of the S Allele of CR1. The analysis of the structure and polymorphisms of the human CR1 gene was performed by restriction mapping of phage and cosmid clones identified by a series of cDNA probes spanning the CR1 gene (Fig. 1 A). When large spans of intervening sequences occurred, genomic libraries were screened with derived genomic probes, e.g., 2.22.2. Few cosmid clones containing inserts corresponding to the homologous regions within the CR1 structural gene were identified, perhaps reflecting a high frequency of recombination leading to the loss of such clones. Restriction mapping of the genomic clones was performed with four enzymes that identified genomic polymorphisms, Bam HI, Eco RI, Hind III, and Sac I.

The alignment of 25 overlapping phage and cosmid clones permitted the mapping of 160 kb that contained all of the S allele. The gene can be divided into nine regions: the 5' region containing untranslated and leader sequences; five regions having exons encoding the LHRs; a short region containing the two 3' SCRs that are not included within an LHR; a segment having exons encoding the transmembrane and cytoplasmic regions; and 3' untranslated sequences (Fig. 2). Sequence analysis of the 5' genomic clone revealed a single exon encoding the signal peptide and part of the 5' untranslated sequences. The 3' boundary of this exon was located at the

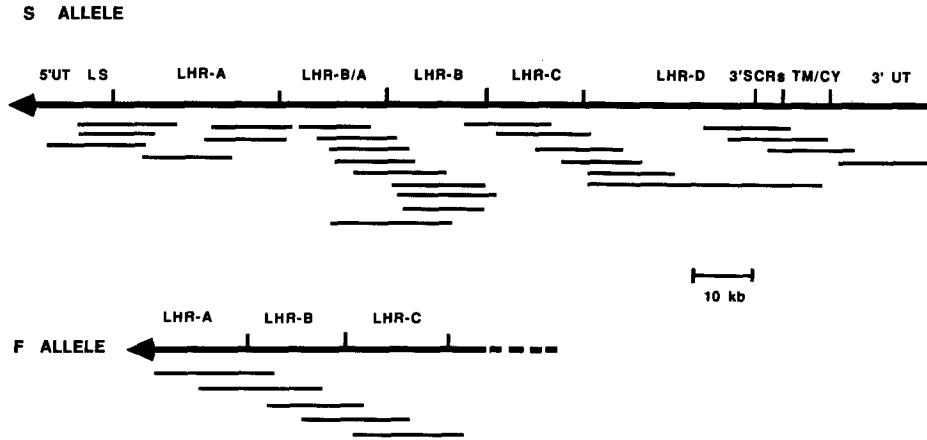


FIGURE 2. The S and F alleles of human CR1. The heavy black bars represent the genes and the lines underneath each represent the overlapping phage or cosmid clones that were aligned to obtain the restriction maps. The vertical marks divide the functional domains of the gene which are designated as in Fig. 1 A.

glycine residue at position 41 (Fig. 3) which is predicted to be the site for the cleavage of the signal peptide (22). At the 3' end of the gene, separate exons encode the last two SCRs as demonstrated by the hybridization of CR1-8 and 1-11 to two distinct Nsi I fragments and the absence of a Nsi I site in the cDNA (Fig. 4 and reference 21). The rest of the 3' coding sequence is contained within three exons, one encoding residues 1969-1976 and the others containing the transmembrane and cytoplasmic domains. The 3' untranslated sequences detected by CR1-15 are located ~9 kb further downstream (Fig. 4).

Between these 5' and 3' genomic regions were the five segments ranging in length from 18 to 30 kb that hybridized to cDNA probes specific for LHR-A, -B, -C, and -D (Figs. 2 and 4). The designation of the genomic segments according to the LHRs uses the nomenclature previously used for the cDNA sequences of the F allotype (21, 22). Despite the high level of similarities between cDNA sequences of different LHRs that resulted in crosshybridization of certain probes under stringent condi-

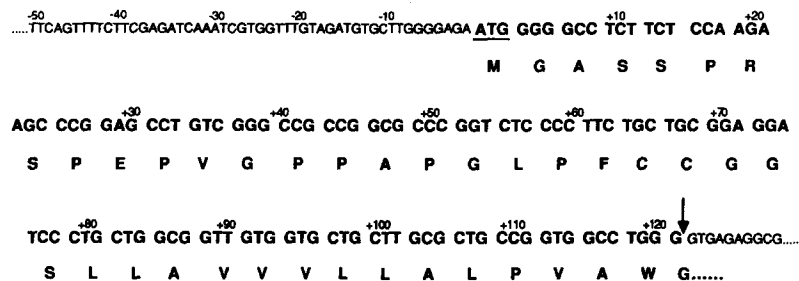


FIGURE 3. Nucleotide and derived amino acid sequences from genomic clone 5.2 showing the intron/exon junction following the leader sequence. Residues -1 to -27 of the 5' untranslated sequence are identical to those found in the cDNA clone T109.4 (22). The arrow designates the proposed splice site followed by the consensus 3' splice sequence.

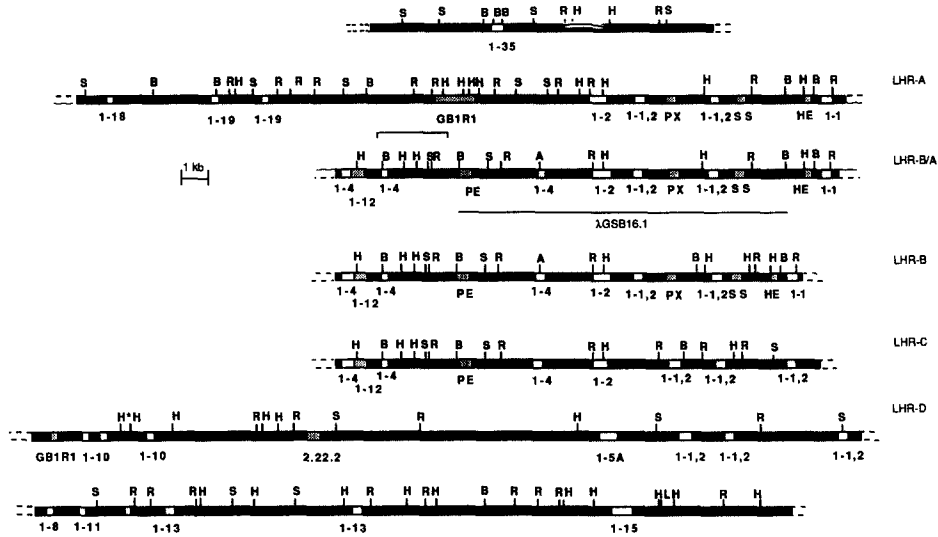


FIGURE 4. Restriction map of the S allele of CR1. The black bars represent contiguous regions of the gene and the white and stippled boxes are the regions to which the cDNA and the genomic probes hybridized, respectively. The placement of the probes on the map was determined by restriction mapping, Southern blot analyses and comparison of these restriction sites to those in the cDNA sequence (21, 22). Some of these placements were approximated and sequence analyses have not been performed in all cases to determine the exact intron/exon boundaries. The bracket over the 5' end of LHR-B/A designates the 2.7-kb gap to which no phage clones could be unambiguously assigned. The restriction sites are designated as follows: A, Ava I; B, Bam HI; H, Hind III; L, Sal I; R, Eco RI, and S, Sac I. Not all Ava I sites are shown. The Hind III site marked with an asterisk in LHR-D corresponds to the polymorphic site described in Figs. 7 and 8.

tions, each LHR genomic region could be distinguished either because of unique hybridizations with cDNA probes or because of distinct restriction maps (Fig. 4). The LHR-A genomic region was identified by its unique hybridization with the CR1-18 and CR1-19 probes, whereas recognition of the LHR-B and -C genomic segments was possible because their 3' regions differentially hybridized with the intervening sequence probes PX, SS, and HE. The LHR-D region could be distinguished by its hybridization with CR1-10 and CR1-5A.

The genomic region present between the LHR-A and -B segments contained a 14.5-kb Bam HI fragment and a 19-kb Sac I fragment that were previously found to be associated with expression of the S allotype. This region most resembled a hybrid comprising the 5' half of the LHR-B segment, which can be distinguished from this region of the LHR-C segment by the Ava I site (Fig. 4), and the 3' half of the LHR-A segment and is therefore termed LHR-B/A. It is considered to contain exons encoding the fifth LHR that had been postulated to be present in the S allotype (19). At the 5' end of the LHR-B/A genomic segment, there is an apparent gap of 2.7 kb to which no phage clones could be assigned. However, when Southern blots containing the Bam HI, Eco RI, Hind III, and Sac I digests of DNA from individuals homozygous or heterozygous for the F and S allotypes were probed with CR1-4 and PE, identical patterns were observed for the relevant junction regions (Fig. 5).

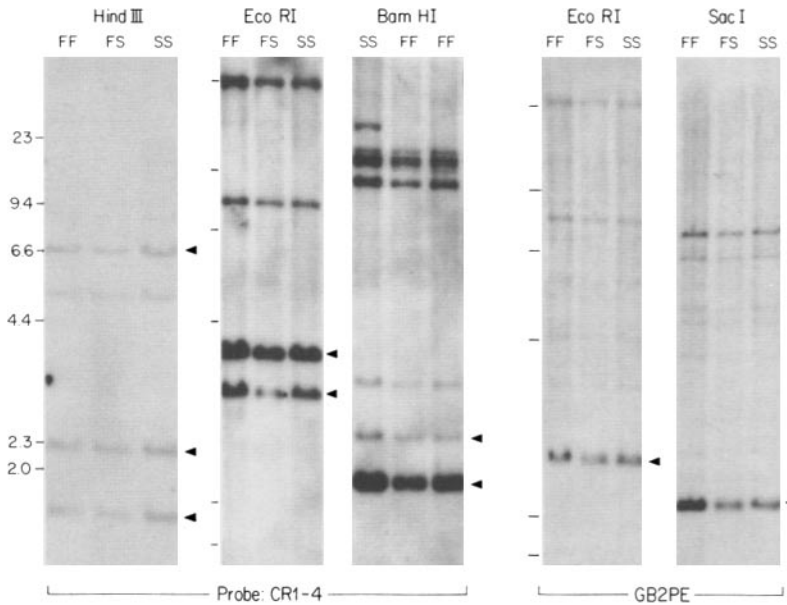


FIGURE 5. Autoradiographs of the Southern blots in which CR1-4 and GB2PE were hybridized to the restriction digests of the DNA from donors expressing the different structural allotypes of CR1. A Hind III digest of the λ DNA was used as a marker and the positions of these fragments are designated in kilobases on the left. The fragments derived from the junctions of LHR-segments in the structural gene are designated by arrows on the right. Additional hybridizing bands were derived from the nonoverlapping cluster described in Fig. 9.

Therefore, the restriction fragments spanning the junctions between the segments of LHR-A and -B, LHR-A, and -B/A, and LHR-B/A and -B were identical in size (Figs. 4 and 6), suggesting that the 2.7-kb gap exists in our map only because of an inability to make a unique assignment of a genomic clone to this region.

Comparison of the LHR segments of the CR1 gene revealed several remarkable similarities. For example, the PE genomic probe from the 5' region of the LHR-B/A segment crosshybridized under stringent conditions to the corresponding regions of the LHR-B and -C genomic segments (Fig. 4) and each site of hybridization was flanked by a series of nine identical restriction sites (Fig. 4). Similarly, the 3' regions of LHR-A and -B crosshybridized with genomic probes from the 3' region of the LHR-B/A segment and these regions also contained similar restriction sites.

In contrast to these highly homologous regions of the CR1 gene were unique segments that included the 5' portion of LHR-A, the 3' portion of LHR-C, and all of LHR-D (Fig. 4). Although the intervening sequence probe GB1R1 hybridized to the 5' introns in both the LHR-A and -D segments, these regions did not have a similar relationship to flanking exons. The intron recognized by this probe in the LHR-A segment was between exons encoding SCR-2 and -3, whereas in the LHR-D segment, the intron was 5' to the exons encoding the first two SCRs of this LHR (Fig. 4). Differences could be found even in genomic segments containing exons having almost identical sequences. For example, although SCR-3 through -7 of LHR-A were 98% identical in nucleotide sequence to SCR-10 through -14 of LHR-B, the lat-

ter could be distinguished at the genomic level by additional Bam HI and Hind III sites and the hybridization of the SS and HE genomic probes to smaller 1.8-kb Hind III and 1.5 kb Eco RI fragments, respectively (Fig. 4). These differences highlight the similarities between the LHR-B/A genomic segment and the homologous regions of the LHR-A and -B segments and strongly suggest that the coding sequences of LHR-B/A will be similar respectively to LHR-B and -A (Fig. 4).

The intron/exon organization of the five homologous segments of the CR1 gene could be inferred from the restriction map and the pattern of hybridization of cDNA probes. In the LHR-A region, the hybridization of cDNA probes specific for SCR-1 and -2 to three distinct Bam HI fragments (Fig. 4) and the finding of noncoding sequences between the codons for residues 134 and 135 of SCR-2 (Bartow, T. J., unpublished observations) indicate that SCR-2 is encoded by at least two exons. Similarly, CR1-4, a 381-bp cDNA probe encoding SCR-15 and SCR-16 of LHR-C that is highly homologous to the corresponding two SCRs of LHR-B/A and -B, hybridized to three Hind III fragments of 7.3, 2.2, and 1.8 kb on Southern blots of DNA from individuals having the FF, FS, and SS allotypes (Fig. 5). As the restriction maps of these respective regions of the CR1 gene were identical (Fig. 4) and there were no Hind III sites within CR1-4, the occurrence of three hybridizing fragments indicates that at least three exons encode each pair of NH₂-terminal SCRs of these LHRs. Furthermore, CR1-12 is a cDNA probe derived from unspliced intervening sequence between residues 551 and 552 of SCR-8 and -9, respectively (21). It hybridized to the 1.8-kb Hind III fragment from each LHR region, suggesting that the first SCR of each LHR is probably encoded by a single exon. The second SCR of LHR-B/A, -B, and LHR-C must be encoded by at least two exons separated by the intervening sequence identified by the probe, PE (Fig. 4). In LHR-D, intervening sequences have been found between three distinct regions that hybridized to CR1-10, the cDNA probe encoding the NH₂-terminal two SCRs of LHR-D (Figs. 4 and 8). Thus, the second SCR of each LHR is encoded by at least two exons.

The Eco RI-Hind III fragment of ~250 bp, which was conserved among LHRs-A, -B/A, -B, and -C, may contain the third and fourth SCRs since both enzyme sites were present in the cDNA and no hybridization to CR1 or CR1-2 was observed in the DNA fragments that were 5' of this region in each LHR. Thus, these two SCRs are encoded either by a single exon or by two exons separated by a small intron. The location of the intervening sequence probes among regions of cDNA hybridization indicates that at least four exons encode the third to the seventh SCR of LHR-A, -B/A, and -B. Similarly, in LHR-C and -D, the CR1-1 and CR1-2 probes hybridized to four distinct fragments (Fig. 4), and sequence analysis showed that the seventh SCR in LHR-C was contained within one exon (21). Thus, at least seven exons are in each LHR, and combining these with at least two exons encoding the two -COOH-terminal SCRs not contained within an LHR (Fig. 4), adds up to at least 37 exons encoding the ecto-domain of the S allotype of CR1.

The Organization of the F Allele of CR1. To obtain further evidence that the F allotype of CR1, which contains four LHRs, is not the product of alternative splicing of a transcript encoding five LHRs, genomic segments of the three 5' LHRs of an F allele were mapped by aligning genomic clones from a library constructed with DNA from an FF individual. The hybridization of GB1R1 in the most 5' clones localized the LHR-A genomic region and the restriction sites in this region correspond to those of the LHR-A segment of the S allele (Figs. 4 and 6). Downstream of this

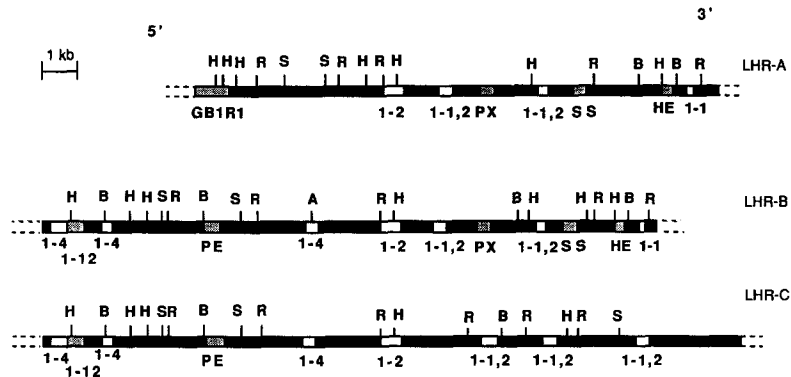


FIGURE 6. Partial restriction map of LHR-A, -B, and -C of the F allele. The black bars represent contiguous segments of the gene while the white and stippled boxes are the regions to which the cDNA and the genomic probes hybridized, respectively. The restriction sites are designated as in Fig. 4.

region was a segment distinguished by Hind III and Eco RI restriction fragments of 1.8 and 1.5 kb, respectively, that hybridized to the genomic probes, SS and HE, making this region identical to the LHR-B rather than the LHR-B/A region of the S allele. The third homologous region was identical to the LHR-C segment and lacked hybridization to the intervening sequence probes, PX, SS, and HE (Figs. 4 and 6). Although the LHR-D region was not included in the clones analyzed, its presence in the F allele is demonstrated by the cDNA sequence (21) and from analysis by Southern blots of DNA from FF homozygotes. In summary, the F allele does not contain the LHR-B/A genomic region of the S allele, indicating that its primary transcript contains four rather than five LHRs.

Definition of the Hind III Polymorphism Associated with the CR1 Quantitative Polymorphism. The allelic Hind III fragments of 7.4 and 6.9 kb that correlated with high and low expression, respectively, of CR1 on erythrocytes were localized to the LHR-D genomic region by their hybridization to the CR1-10 cDNA probe (Fig. 7). In this Southern

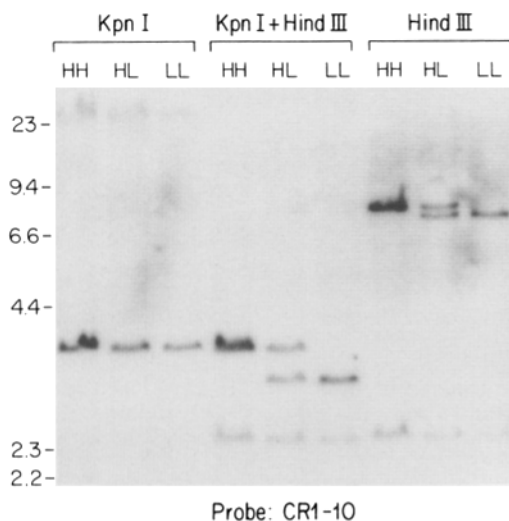


FIGURE 7. Autoradiograph of the Southern blot in which CR1-10 was hybridized to the Hind III and/or Kpn I digests of the DNA of donors expressing high (HH), intermediate (HL), or low (LL) numbers of CR1 on their erythrocytes. A Hind III digest of λ DNA was used as a marker and the positions of these fragments are designated in kilobases on the left. The fragments derived from the structural gene are designated by arrows on the right.

blot, a fragment of 2.35 kb hybridizing to 1-10 was observed that has not been found in any CR1 genomic clones. Restriction maps of two allelic genomic clones, 2.22 and 2.38, that hybridized with CR1-10 revealed that although they contained identical 3.5-kb Kpn I fragments, double digestion with Kpn I and Hind III generated distinct fragments of 3.0 and 3.5 kb, respectively (Figs. 7 and 8). The additional Hind III site in 2.22 accounting for the smaller fragment was 0.5 kb upstream from the 3' Hind III site of the 7.4-kb restriction fragment in clone 2.38, and therefore corresponded to the site responsible for the 6.9-kb Hind III fragment defining the L allele. Sequence analysis of 126-bp Eco RV to Rsa I fragments from both genomic clones identified the single base change responsible for the additional Hind III site in 2.22 relative to 2.38 and confirmed its location in noncoding sequences of the gene (Fig. 8).

Duplication of the CR1 Gene. Evidence for duplication of the CR1 gene was provided by the finding of a cluster of overlapping phage clones spanning 60 kb that hybridized at high stringency to CR1 cDNA and genomic probes (Fig. 9). The first 36 kb of this cluster contained regions homologous to the signal peptide and the SCRs of LHR-A. Hybridization to the noncoding probes, GB1R1, PX, SS, and HE were also observed in the same 5' to 3' order as in the LHR-A genomic segment. The

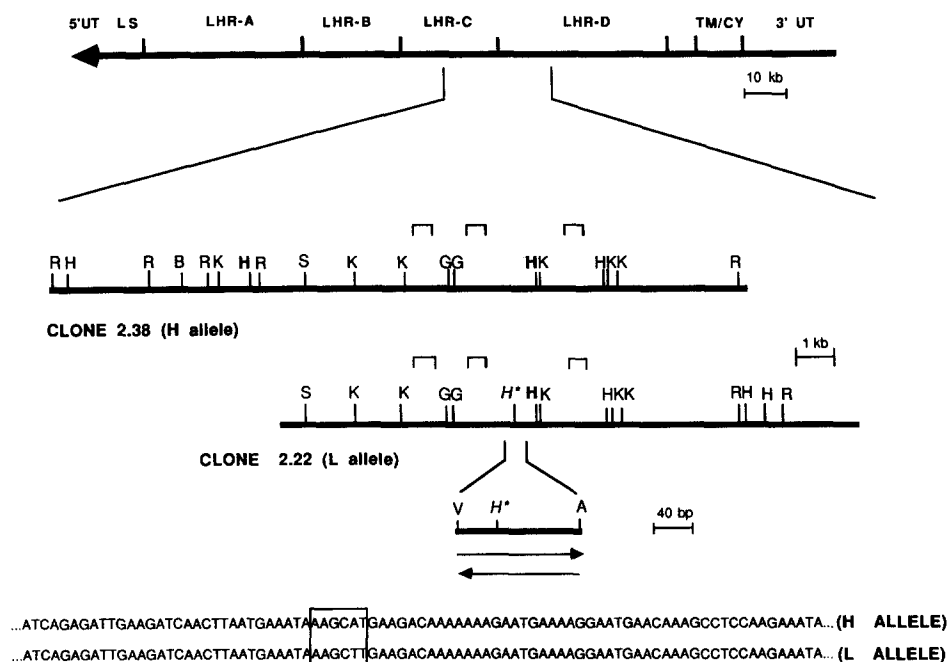


FIGURE 8. Restriction maps, sequence strategy and partial sequence of the inserts of the phage clones 2.38 and 2.22 which contain the allelic 7.4- and 6.9-kb Hind III fragments that are associated with the high (H) or low (L) expression of CR1, respectively. The restriction sites are as follows: A, Rsa I; B, Bam HI; G, Bgl I; H, Hind III; K, Kpn I; R, Eco RI; S, Sac I and V, Eco RV. The polymorphic Hind III site is designated in italics with an asterisk and the Hind III sites that flank the 7.4- and 6.9-kb fragments are in boldface. The brackets above the clones represent the regions which hybridize to the CR1-10 probe. The bases corresponding to the new Hind III site are enclosed in a box.

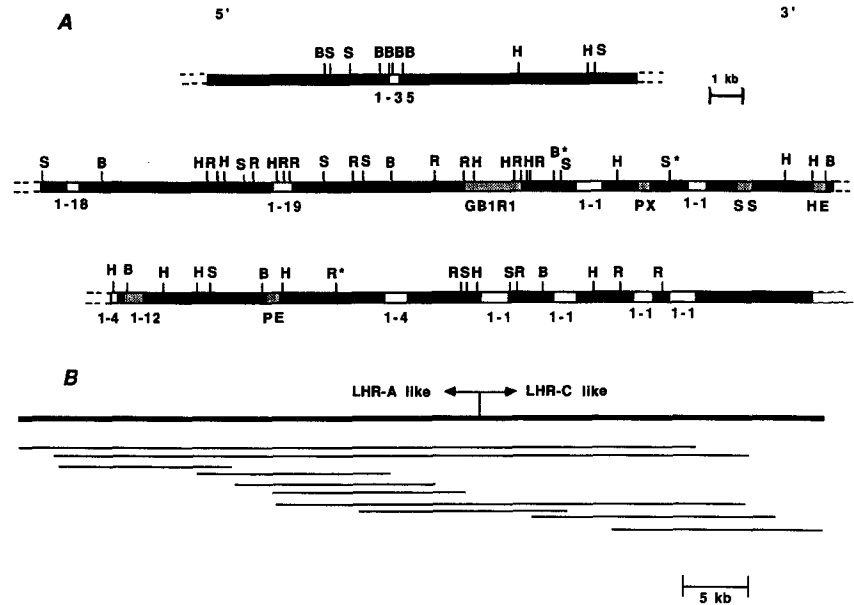


FIGURE 9. (A) Restriction map of the CR1-like nonoverlapping cluster of genomic clones. The black bars represent the gene and the stippled boxes are the regions to which the cDNA and the genomic probes hybridized, respectively. The restriction sites are designated as in Fig. 4. The polymorphic sites are designated with asterisks. (B) The heavy line represents the gene and the lighter lines represent the overlapping phage or cosmid clones used to generate the restriction map in (A).

3' portion of this segment of DNA resembled the LHR-C more than the LHR-B region in its capacity to hybridize to the CR1-4, 1-12, PE, and 1-1 probes and its lack of hybridization to the PX, SS, and HE probes. However, the restriction map for these clones was clearly distinct from that for the CR1 structural gene, indicating that this is a duplicated and not allelic region (Fig. 9).

Three polymorphic sites that accounted for previously identified genomic polymorphisms have been mapped to this region. The absence or presence of an Eco RI site located 16 kb from the 3' end of this cluster determines the allelic 20- and 16-kb Eco RI fragments hybridizing with CR1-4 and CR1-4 that have been found in DNA from all 85 individuals that have been examined (19). The absence of a Bam HI site 28 kb from the 5' end of this second cluster determines the appearance of a 14.5-kb Bam HI fragment hybridizing with GB1R1 that corresponds to the insert of clone 4.2 (see Materials and Methods) in ~10% of normal individuals (19). This Bam HI polymorphism is associated with the presence of an additional Sac I site 3.5 kb downstream that accounts for the appearance of 7.9- and 3.3-kb fragments that hybridize with CR1-4 and CR1-1, respectively (19). Whether these polymorphisms have any effects on CR1 expression remains unknown since linkage disequilibrium has not been observed with the structural or quantitative polymorphisms (19).

Discussion

The S allele of human CR1 contained five LHR segments of 20 to 30 kb each, while the cDNA and a partial genomic map of the F allele showed that there were

only four homologous regions (Figs. 1, 2, 4, and 6). The longer transcripts associated with the S allotype were products of this longer allele with coding capacity for an additional seven SCRs or ~ 450 aa. Although sequence homology existed in both the coding and noncoding regions, each LHR segment had unique restriction sites or introns. The association of the 14.5-kb Bam HI and the 19-kb Sac I fragments with the LHR-B/A segment identified this as the additional DNA in the S allele (Figs. 4 and 6). The similarity of this segment to the 5' half of the LHR-B and the 3' half of the LHR-A segments predicts that its coding sequence will be identical to that of LHR-B. As LHR-B and -C have recently been shown to contain C3b-binding sites (22), the S allotype of CR1 may have a third binding site for this ligand.

In the CR1 gene, the signal peptide and the transmembrane and cytoplasmic regions are encoded by separate exons, while most of the SCRs are found in exons of ~ 200 bp as has been reported in human C4BP, C2, factor B and IL-2R (Tac-I) and in murine factor H (34–37). That the second SCR in each LHR is encoded by two exons is an indication of specialized functional capacities of this repeat unit. Recent evidence from the analyses of the C3b/C4b binding capacities of CR1 deletion mutants lacking one or more LHRs indicated that the specificity for these ligands is determined by the first two SCRs in each LHR (22). These observations are consistent with the second SCR in murine factor H being encoded by two exons and the localization of the C3b binding site of human factor H to the NH₂ terminus of this molecule (37, 38).

Within the CR1 gene, the regions of homology or divergence at the genomic level corresponded to those found in the cDNA, suggesting that each LHR evolved as a 20–30-kb genomic unit. For example, as in the cDNA, the genomic regions that contain the first two SCRs of LHR-B and -C were almost identical (Figs. 1, 4, and 6). Thus, the divergence of LHR-A in the first two SCRs and their flanking introns presumably reflects its preference for the ligand C4b over C3b (22). Conservation of the third and fourth SCRs in LHR-A, -B/A, -B, and -C was observed at both the cDNA and the genomic level, indicating a contribution of these units either to the functional capacities of their neighboring SCRs or to other necessities for structural constraints. In contrast to the similarities in both introns and exons between the LHR-A and -B segments, only the preservation of coding sequences was observed in the LHR-C and -D genomic segments (Fig. 4 and reference 21).

The extended internal homologies observed in the elongated structure of CR1 provide opportunities for homologous recombination between two alleles in which the generation of the different allotypes of CR1 can be achieved by the addition or deletion of one or more LHRs. For example, acquisition of LHR-B/A as a hybrid of the 5' half of LHR-B and the 3' half of LHR-A can be achieved by a single crossover at the midpoints of LHR-A and -B between two F alleles. The reciprocal result of such an event might be an allele that lacks the 3' portion of LHR-A and the 5' portion of LHR-B and contains only one C3b binding site (Fig. 10). The F' or C allotype that is ~ 40 kD smaller than the F allotype may be this reciprocal product of the recombination event that generated the S allotype. The presumed monovalency for C3b of the F' allotype might impair binding of C3b-bearing immune complexes, accounting for the association of this allotype with systemic lupus erythematosus (16, 39). A single crossover at the midpoints of LHR-A and -C of two F alleles can also account for an allele which contains six LHRs (17) and perhaps four C3b binding

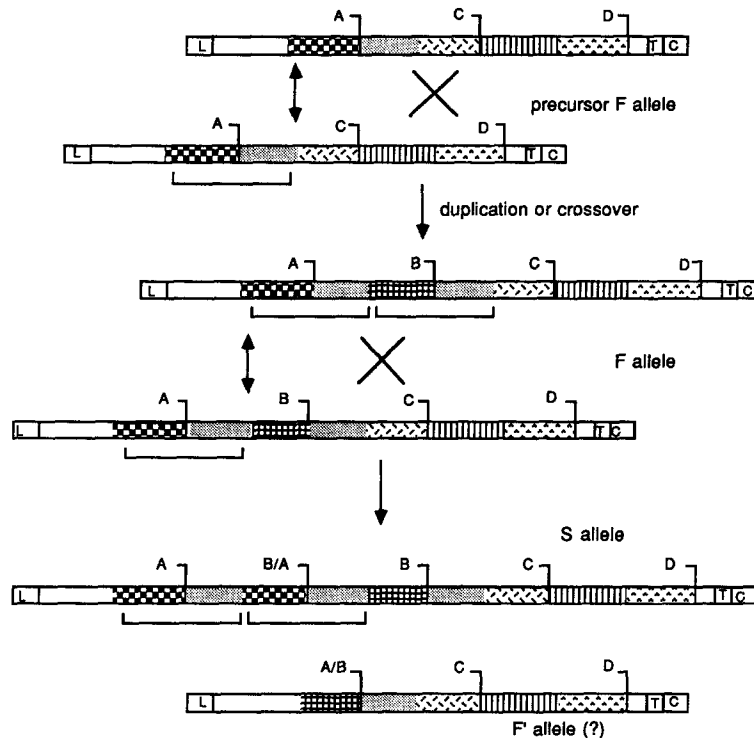


FIGURE 10. Proposed scheme for the generation of the CR1 allotypes. The ancestral gene presumably contained a signal peptide (L), three LHR segments, a transmembrane (T) and a cytoplasmic domain (C), designated by separate boxes. The different patterns in the LHRs represent regions of divergence. The double arrows designate the points of crossover and the brackets under the 3' end of LHR-A and the 5' end of LHR-C or -B represent the segments of DNA that have been duplicated to generate the F and the S alleles, respectively.

sites. Similarly, the presentation of LHR-B as a hybrid of the 5' half of LHR-C and the 3' half of LHR-A at both the cDNA and the genomic level suggests that it can also be acquired by a similar crossover between two F allele precursors containing three LHRs (Fig. 10). An alternate mechanism for the insertion of an LHR involves direct duplication, and conversion events are necessary to maintain the high homology among both introns and exons. These duplications within CR1 may be driven by the combined necessity to extend the ligand binding sites away from the cell surface and the requirement to achieve increased avidity for C3b/C4b-coated surfaces through multivalent interactions.

Duplicative events were probably responsible for the generation of functionally and structurally related genes within the RCA region (7-11). The finding of an additional cluster of clones with CR1 related sequences provides further evidence for CR1 duplication (Fig. 9). The close proximity of this second cluster of clones has been inferred from pedigree analyses in which the the Eco RI polymorphism which mapped to the second gene was found to be linked to the structural polymorphism of CR1 (19). Furthermore, in the analyses of the RCA region by pulse-field electrophoresis, hybridization to repetitive or 5' and 3' CR1 cDNA probes was observed within a single restriction fragment of ~250 kb (12, 13).

An estimation of the sequences which are homologous to the cDNA probes CR1-1 and CR1-4 indicates that the nonoverlapping cluster of clones contains coding capacity for two LHRs. However, when these probes were used in Northern blot analyses, no transcriptional products other than those relating to CR1 were identified in tonsils, peripheral blood leukocytes, and the myelomonocytic cell line, HL60, indicating that the additional gene may not be transcriptionally active in hematopoietic cells (19, 20, 40, 41). Furthermore, the finding of three genomic polymorphisms within 16 kb of this cluster suggests that mutations may be occurring more frequently relative to the structural gene. These clones may therefore represent a partial allele generated during the multiple duplicative processes in the evolution of CR1 or part of a pre-CR1 gene having only LHR-A, -C, and -D, as depicted in Fig. 10. Evidence in support of this hypothesis is the identification of an additional CR1-10-hybridizing Hind III fragment of 2.35 kb that has not been localized to the structural gene (Fig. 7), suggesting that another segment similar to LHR-D may exist.

The Hind III polymorphism that was associated with the high or low expression of CR1 on erythrocytes was localized to a single base change in the intervening sequences between the exons encoding the second SCR in LHR-D. The occurrence of this polymorphism 35 kb away from the 3' end of the LHR-B/A region accounts for the lack of recombination between these genomic polymorphisms (19). Although intragenic enhancer elements have been reported for the Ig genes, these regulatory sequences are within several kilobases of the promoter region (reviewed in reference 42). Since the additional Hind III site is at least 70 kb away from the 5' promoter region in an F allele, the effect of this base change on the expression of CR1 remains unclear. However, the finding of individuals who were homozygous for the 7.4-kb Hind III fragment associated with high expression of CR1 but who consistently expressed low levels of CR1 (25) indicated that other genetic factors may be involved and the Hind III site may represent a genetic marker for other *cis* factor(s) that regulate the expression of CR1.

Summary

Structural and quantitative polymorphisms have been described in human CR1. In the former, the S allotype is larger than the F allotype by 40–50 kD, the size of a long homologous repeat (LHR). In the latter, homozygotes for a 7.4-kb Hind III fragment express fourfold more CR1 per erythrocyte than do homozygotes for the allelic 6.9-kb restriction fragment. The basis for these genomic polymorphisms has been determined by restriction mapping the entire S allele and part of the F allele. The S allele is 158 kb and contains 5 LHRs of 20–30 kb, designated -A, -B/A, -B, -C, and -D, respectively, 5' to 3'. Extensive homology was found among the LHRs in their restriction maps, exon organization, and the coding and noncoding sequences. The presence of LHR-B/A in the S allele but not in the F allele accounts for the longer transcripts and polypeptide associated with the former allotype. At least 42 exons are present in the S allele, with distinct exons for the leader sequence, the transmembrane and cytoplasmic regions and most of the SCRs comprising the extracellular portion of CR1. Consistent with the mapping of the ligand binding site to the first two SCRs in each LHR, the second SCRs in LHR-A, -B/A, -B, and -C are encoded by two exons, reflecting a specialized function for this unit. The allelic 7.4/6.9-kb Hind III fragments extend from the 3' region of LHR-C to LHR-D. The

6.9-kb restriction fragment is the result of a new Hind III site generated by a single base change in the intron between the exons encoding the second SCR of LHR-D. A second cluster of genomic clones has been identified by hybridization to CR1 probes. Although they contain regions of hybridization to the cDNA and genomic probes derived from CR1, these cannot be overlapped with the structural gene owing to their distinct restriction maps. Three genomic polymorphisms previously identified by CR1 cDNA probes map to this region. These additional clones may represent part of a duplicated allele located nearby within the CR1 locus.

The authors would like to thank Dr. Thomas J. Bartow for providing the nucleotide sequences which determined the 3' exon/intron junction within SCR-2.

Received for publication 18 August 1988 and in revised form 17 November 1988.

References

1. Fearon, D. T., and W. W. Wong. 1983. Complement ligand-receptor interactions that mediate biologic responses. *Annu. Rev. Immunol.* 1:243.
2. Reynes, M., J. P. Aubert, J. H. M. Cohen, J. Audouin, V. Tricotet, J. Diebold, and M. D. Kazatchkine. 1985. Human follicular cells express CR1, CR2, and CR3 complement receptor antigens. *J. Immunol.* 135:2687.
3. Yoon, S. H., and D. T. Fearon. 1985. Characterization of a soluble form of the C3b/C4b receptor (CR1) in human plasma. *J. Immunol.* 134:3332.
4. Cornacoff, J. B., L. A. Hebert, W. L. Smead, M. E. Van Aman, D. J. Birmingham, and F. J. Waxman. 1983. Primate erythrocyte-immune complex-clearing mechanism. *J. Clin. Invest.* 71:236.
5. Fearon, D. T. 1979. Regulation of the amplification C3 convertase of human complement by an inhibitory protein isolated from human erythrocyte membrane. *Proc. Natl. Acad. Sci. USA.* 76:5867.
6. Iida, K., and V. Nussenzweig. 1981. Complement receptor is an inhibitor of the complement cascade. *J. Exp. Med.* 153:1138.
7. Rodriguez de Cordoba, S., D. M. Lublin, P. Rubinstein, and J. P. Atkinson. 1985. Human genes for the three complement components that regulate the activation of C3 are tightly linked. *J. Exp. Med.* 161:1189.
8. Weis, J. H., C. C. Morton, G. A. P. Bruns, J. J. Weis, L. B. Klickstein, W. W. Wong, and D. T. Fearon. 1987. Definition of a complement receptor locus: the genes encoding the C3b receptor and the C3d/Epstein-Barr virus receptor map to human chromosome 1. *J. Immunol.* 138:312.
9. Rey-Campos, J., P. Rubinstein, and S. Rodriguez de Cordoba. 1987. Decay-accelerating factor: genetic polymorphism and linkage to the RCA (regulator of complement activation) gene cluster in humans. *J. Exp. Med.* 166:246.
10. Lublin, D. M., R. S. Lemons, M. M. Le Beau, V. M. Holers, M. L. Tykocinski, M. E. Medof, and J. P. Atkinson. 1987. The gene encoding decay-accelerating factor (DAF) is located in the complement-regulatory locus on the long arm of chromosome 1. *J. Exp. Med.* 165:1731.
11. Lublin, D. M., M. K. Liszewski, T. W. Post, M. A. Arce, M. M. Le Beau, R. S. Lemons, T. Seya, and J. P. Atkinson. 1988. Molecular cloning and chromosomal localization of human complement membrane cofactor protein (MCP). *J. Exp. Med.* 168:181.
12. Rey-Campos, J., P. Rubinstein, and S. Rodriguez de Cordoba. 1988. A physical map of the human regulator of complement activation gene cluster linking the complement genes CR1, CR2, DAF and C4BP. *J. Exp. Med.* 167:664.

13. Carroll, M. C., E. M. Alicot, P. J. Katzman, L. B. Klickstein, J. A. Smith, and D. T. Fearon. 1988. Organization of the genes encoding complement receptors type 1 and 2, decay-accelerating factor, and C4-binding protein in the RCA locus on human chromosome 1. *J. Exp. Med.* 167:1271.
14. Wong, W. W., J. G. Wilson, and D. R. Fearon. 1983. Genetic regulation of a structural polymorphism of human C3b receptor. *J. Clin. Invest.* 72:685.
15. Dykman, T. R., J. L. Cole, K. Iida, and J. P. Atkinson. 1983. Polymorphism of human erythrocyte C3b/C4b receptor. *Proc. Natl. Acad. Sci. USA.* 80:1698.
16. Dykman, T. R., J. A. Hatch, and J. P. Atkinson. 1984. Polymorphism of the human C3b/C4b receptor. Identification of a third allele and analysis of receptor phenotypes in families and patients with systemic lupus erythematosus. *J. Exp. Med.* 159:691.
17. Dykman, T. R., J. A. Hatch, M. S. Aqua, and J. P. Atkinson. 1985. Polymorphism of the C3b/C4b (CR1) receptor: characterization of a fourth allele. *J. Immunol.* 134:1787.
18. Lublin, D. M., R. C. Griffith, J. P. Atkinson. 1986. Influence of glycosylation on allelic and cell specific *M_r* variation, receptor processing and ligand binding of the human complement C3b/C4b receptor. *J. Biol. Chem.* 261:5736.
19. Wong, W. W., C. A. Kennedy, E. T. Bonaccio, J. G. Wilson, L. B. Klickstein, J. H. Weis, and D. T. Fearon. 1986. Analysis of multiple restriction fragment length polymorphisms of the gene for the human complement receptor type I. Duplication of genomic sequences occurs in association with a high molecular mass receptor allotype. *J. Exp. Med.* 164:1531.
20. Holers, V. M., D. D. Chaplin, J. F. Leykam, B. A. Gruner, V. Kumar, and J. P. Atkinson. 1987. Human CR1 mRNA polymorphism correlates with the CR1 allelic molecular weight polymorphism. *Proc. Natl. Acad. Sci. USA.* 84:2459.
21. Klickstein, L. B., W. W. Wong, J. A. Smith, J. H. Weis, J. G. Wilson, and D. T. Fearon. 1987. Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristic of C3/C4 binding proteins. *J. Exp. Med.* 165:1095.
22. Klickstein, L. B., T. J. Bartow, V. Miletic, L. D. Rabson, J. A. Smith, and D. T. Fearon. 1988. Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis. *J. Exp. Med.* 168:1699.
23. Wilson, J. G., W. W. Wong, P. H. Schur, and D. T. Fearon. 1982. Mode of inheritance of decreased C3b receptors on erythrocytes of patients with systemic lupus erythematosus. *N. Engl. J. Med.* 307:981.
24. Rodriguez de Cordoba, S., and P. Rubinstein. 1987. Quantitative variations of CR1 in human erythrocytes are controlled by genes within the RCA cluster. *J. Exp. Med.* 164:1274.
25. Wilson, J. G., E. E. Murphy, W. W. Wong, L. B. Klickstein, J. H. Weis, and D. T. Fearon. 1986. Identification of a restriction fragment length polymorphism by a CR1 cDNA that correlates with the number of CR1 on erythrocytes. *J. Exp. Med.* 164:50.
26. Moldenhauer, F., J. David, A. H. L. Fielder, P. J. Lachmann, and M. J. Walport. 1988. Inherited deficiency of erythrocyte complement receptor type 1 does not cause susceptibility to systemic lupus erythematosus. *Arthritis Rheum.* 30:961.
27. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
28. Grosveld, F. G., T. Lund, E. J. Murray, A. L. Mellor, H. H. M. Dahl, and R. A. Flavell. The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Res.* 10:6715.
29. Frischauf, A. M., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170:827.
30. Rackwitz, H.-R., G. Zehetner, A.-M. Frischauf, and H. Lehrach. 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. *Gene.* 30:195.

31. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 75:4853.
32. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 84:4767.
33. Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387.
34. Lintin, S. J., A. Lewin, and K. B. M. Reid. 1988. Studies on the structure of the human C4b-binding protein gene and characterization of the protein signal sequence. *Complement.* 4:186. (Abstr.)
35. Campbell, R. D., and D. R. Bentley. 1985. The structure and genetics of the C2 and factor B genes. *Immunol. Rev.* 87:19.
36. Leonard, W. J., J. M. Depper, M. Kanehisa, M. Kronke, N. J. Pfeffer, P. B. Svetlik, M. Sullivan, and W. C. Greene. 1985. Structure of the interleukin-2 receptor gene. *Science (Wash. DC).* 230:633.
37. Vik, D. P., J. B. Keeney, P. Munoz-Canoves, D. D. Chaplin, and B. F. Tack. 1988. Structure of the murine complement factor H gene. *J. Biol. Chem.* 263:16720.
38. Alsenz, J., T. F. Schultz, J. D. Lambris, R. B. Sim, and M. P. Dierich. 1985. Structural and functional analysis of the complement factor H with the use of different enzymes and monoclonal antibodies to factor H. *Biochem. J.* 232:841.
39. Van Dyne, S., V. M. Holers, D. M. Lublin, and J. P. Atkinson. 1987. The polymorphism of the C3b/C4b receptor in the normal population and in patients with systemic lupus erythematosus. *Clin. Exp. Immunol.* 68:570.
40. Wong, W. W., L. B. Klickstein, J. A. Smith, J. H. Weis, and D. T. Fearon. 1985. Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b. *Proc. Natl. Acad. Sci. USA.* 82:7711.
41. Jack, R. M., and D. T. Fearon. 1988. Selective synthesis of mRNA and proteins by human peripheral blood neutrophils. *J. Immunol.* 140:4286.
42. Gregor, P. D., B. J. Korbin, C. Milcarek, and S. L. Morrison. 1986. Sequences 3' of immunoglobulin heavy chain genes influence their expression. *Immunol. Rev.* 89:32.